

EXPLORING SELECTIVE IMAGE MATCHING METHODS FOR ZERO-SHOT AND FEW-SAMPLE UNSUPERVISED DOMAIN ADAPTATION OF URBAN CANOPY PREDICTION

John Francis

The Alan Turing Institute
London, United Kingdom
jfrancis@turing.ac.uk

Stephen Law

Department of Geography
University College London
London, United Kingdom
stephen.law@ucl.ac.uk

ABSTRACT

We explore simple methods for adapting a trained multi-task UNet which predicts canopy cover and height to a new geographic setting using remotely sensed data without the need of training a domain-adaptive classifier and extensive fine-tuning. Extending previous research, we followed a selective alignment process to identify similar images in the two geographical domains and then tested an array of data-based unsupervised domain adaptation approaches in a zero-shot setting as well as with a small amount of fine-tuning. We find that the selective aligned data-based image matching methods produce promising results in a zero-shot setting, and even more so with a small amount of fine-tuning. These methods outperform both an untransformed baseline and a popular data-based image-to-image translation model. The best performing methods were pixel distribution adaptation and fourier domain adaptation on the canopy cover and height tasks respectively.

1 BACKGROUND

Machine learning algorithms utilizing remote sensing data have been able to estimate numerous features of the environment such as building footprints (Zhu et al., 2021), land cover types (Vali et al., 2020), and urban tree canopy characteristics (Francis et al., 2023). While these algorithms are often able to produce reliable estimates for a particular geographical area, when images from a new environment, or images captured by different instruments at potentially different resolutions are introduced, the reliability of these algorithms falters (Tuia et al., 2016). Additionally, as computer vision models are shared more broadly, either through community driven packages (e.g., <https://sci.vision/>), or are integrated into web tools which allow users to upload their own images for inference (e.g., <https://segment-anything.com/demo>), unsupervised domain adaptation (*UDA*) methods become critical for increasing the utility of these models to new users with their own unique datasets.

In this paper, we examine a multi-task UNet model which simultaneously predicts urban canopy cover as well as canopy height from RGB aerial imagery. This algorithm was trained and evaluated solely using images from Chicago, USA, captured by the National Agriculture Imagery Program NAIP. Here we explore methods for applying this algorithm to a Borough in London, UK, using data collected by the UK’s Ordnance Survey. Instead of fine-tuning or training a domain-adaptive-classifier, our novelty lies in experimenting a variety of simple data-based *UDA* approaches in a zero-shot setting, which doesn’t require any training, or with a small amount of fine-tuning.

2 RELATED WORKS

Unsupervised domain adaptation (*UDA*) techniques leverage models trained in one source domain to perform accurately on a different, unlabeled target domain. These methods can be broadly categorized into model-based and data-based adaptations.

Model-based adaptations modify the underlying source model to improve its performance on the target domain, primarily through adversarial or self-training methods. Ganin et al. (2016) introduced

a domain-classifier that adjoins to the source model, employing a gradient reversal layer to adjust the model’s weight for the target domain. Tzeng et al. (2017) generalised this notion, by pre-training an encoder on source images and then adapting the target encoder using an adversarial classifier. Self-training, on the other hand, generates target pseudo labels from models trained from the source domains, which are then used to iteratively refine the model for the target domain (Zou et al., 2018).

In contrast, data-based adaptation focuses on altering images either in the source or target domain so that the two are statistically similar. This can be accomplished through unpaired image-to-image translation models like CycleGAN (Zhu et al., 2017), which suffers from changes in semantic content. More recent efforts such as ColorMapGAN (Tasar et al., 2020) try to tackle this limitation by restricting the model to only perform pixel-wise intensity transformations, achieving promising results. However, both model-based approaches and image-to-image data-based approaches require a separate model to be learned between domain pairs, a challenge (energy and costs considerations) that becomes impractical with an increasing number of domains Strubell et al. (2019).

Instead, this research focuses on more straightforward data-based approaches with simple image-matching methods, whereby no training or modifications are made to the underlying source model or the need for an image translation model. Related works in remote sensing includes Yaras et al. (2023), who designed a random-selective histogram matching algorithm for searching and then aligning similar images from source and target domain. Expanding upon this randomised selective search, we evaluate an array of image matching algorithms beyond histogram matching for tree canopy cover and height prediction.

3 METHOD

Our aim in this research is to test the efficacy of selective-search image-matching methods to align an RGB image $I \in \mathbb{R}^{h \times w \times 3}$ in the new target domain (London) to be used on models trained in the source domain (Chicago) in a zero-shot setting or without extensive fine-tuning and data augmentation. In this work, we transform an image I_{tgt} in the target domain into the same or similar image $I_{tgt \rightarrow src}$ in the source domain by $I_{tgt \rightarrow src} = T_{tgt \rightarrow src}(I_{tgt})$ so that a model trained in the source domain $Y_{src} = f_{src}(I_{src}; \theta)$ parameterised by θ can be used for prediction using the translated target image $Y_{tgt \rightarrow src} = f_{src}(I_{tgt \rightarrow src})$ without retraining or extensive fine-tuning or data augmentation.

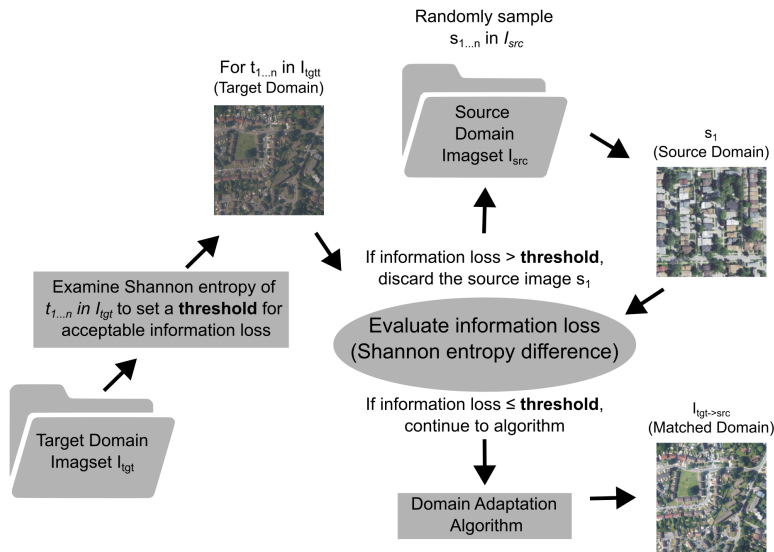


Figure 1: Randomized Image Matching

The data for this study includes 812 1m resolution RGB images from each of London, UK ¹ and Chicago, USA ² covering about 47 square kilometres. London images were chosen based on their alignment with available and temporally aligned LiDAR point cloud data at roughly 12 points per square metre ³ which could be used to generate estimates of ground truth tree canopy cover and height for evaluation, using methods similar to Roussel et al. (2020); Francis et al. (2023). Chicago images were randomly selected to provide a comparable set of source domain images for the domain adaptation algorithms. The London images were originally collected at a 25cm resolution while the Chicago images were collected at a 1m resolution, so prior to any domain adaptation approaches the London images were converted to 1m using a simple mean aggregation.

Figure 1 shows the randomized image matching process we used to align images from London and Chicago. Similar to Yaras et al. (2023), we used the Shannon entropy measure $S = -\sum p(x)\log(p(x))$ to evaluate the likelihood of a paired reference image leading to a large amount of information loss during the translation of an image from the new target domain to the source domain. We randomly sampled images from the source domain (Chicago) until a viable match was found to ensure minimal information loss in the images of the target domain (London), leading to more robust transformations.

We tested four simple data-based image matching approaches, histogram matching in RGB space (Yaras et al., 2023), histogram matching in LAB space, fourier domain adaptation (FDA) (Yang & Soatto, 2020)⁴ and a naive pixel distribution adaptation (PDA) that simple transforms(PCA) an image fitted in the source domain and then inverse transform fitted from the target domain. The transformation were completed using the scikit-image (van der Walt et al., 2014) and Albumentations (Buslaev et al., 2020) python packages.

We additionally tested one data-based image-to-image translation model, CycleGAN Zhu et al. (2017), as well as the target (London) images without transformation, as baselines. We evaluate each of these methods in a zero-shot setting, testing the pre-trained UNet (Francis et al., 2023) in the source domain of Chicago on the transformed images, and then fine-tune separate versions of the algorithm using the transformed datasets. The canopy cover task is evaluated using intersection over union (IoU), while the canopy height task is evaluated using mean absolute error (MAE).

For model fine-tuning, the 812 transformed images in the target domain were split into a train and test set using an 80/20 split. The algorithm was tuned for a maximum of 75 epochs with early stopping if the model failed to improve after 10 epochs. An Adam optimizer with an initial learning rate of 0.001 was used, decaying by a gamma of 0.905 every five epochs. Following Francis et al. (2023), the loss function is a simple weighted additive loss which is a combination of mean squared loss for the canopy height task and binary cross entropy for the tree canopy cover task $L_{total} = \gamma L_{canopy} + \lambda L_{height}$ where we chose the best parameter λ based on previous work in the source domain.

4 RESULTS

Figure 2 shows an example image transformation for each of the methods evaluated in this paper, alongside its ground truth and predicted estimates for tree canopy cover and height. All the data adaptation methods have perceptually altered the pixel intensity of the target image. PDA in particular made the target image slightly brighter and more similar to the source domain. Conversely, CycleGAN has shifted the target image closest to the source domain but the semantics of the image have also been unrealistically altered (e.g. changes in buildings and roads) (Tasar et al., 2020).

Table 1 shows the zero-shot results for the source model evaluated on the transformed target images. For the canopy cover task, PDA produced the best results with a mean IoU of 0.5131. FDA and LAB-space histogram matching both performed worse than evaluating the original target images without any image transformations. Conversely, for the canopy height task, the most performative method was FDA with a MAE of only 0.5864, which is even better than the results of the algorithm

¹<https://digimap.edina.ac.uk/aerial>

²<https://naip-usdaonline.hub.arcgis.com/>

³<https://environment.data.gov.uk/survey>

⁴<https://github.com/YanchaoYang/FDA>

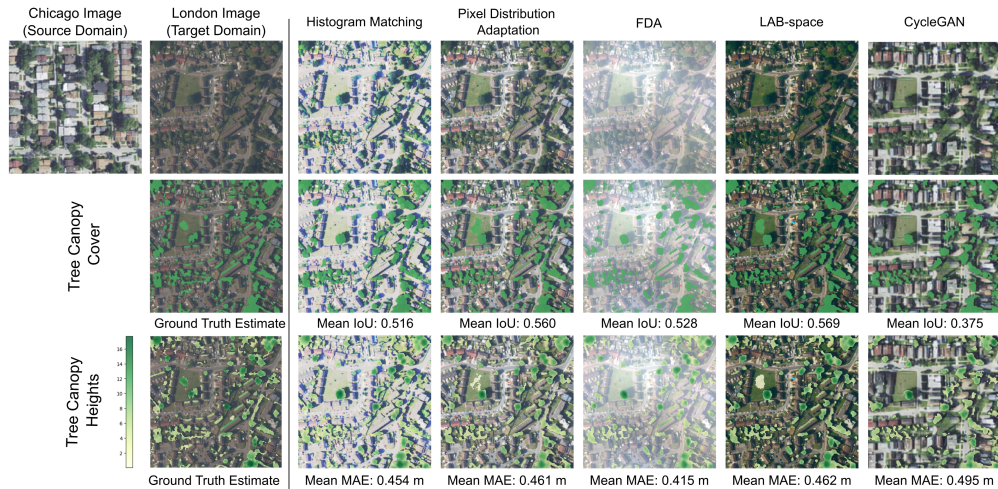


Figure 2: Example outputs from different Unsupervised Domain Adaptation Methods.

on the source domain images. Histogram matching, PDA and CycleGAN all performed worse on the canopy height task than the original target images without any image transformations.

The results of the small samples fine-tuned algorithms can be seen in Table 2. Similar to the zero-shot results, PDA was the best-performing method for the canopy cover task with an IoU of 0.7014. Histogram matching, LAB-space histogram matching, and CycleGAN all performed worse than a model fine-tuned on the un-transformed target images. For the canopy height task, PDA also had the lowest MAE at 0.5547 narrowly edging past LAB-space histogram matching. All of the simple image matching methods outperformed the algorithm fine-tuned on the un-transformed target images for the canopy height task, while CycleGAN performed worse than the un-transformed images on both tasks.

Table 1: Zero-shot Unsupervised Domain Adaptation

Data	Method	mIoU	MAE(m)
Source	NA	0.6650	0.6538
Target	None	0.3709	0.8321
Target	HM	0.4784	1.0302
Target	FDA	0.3659	0.5864
Target	PDA	0.5131	0.9312
Target	LAB-HM	0.3695	<u>0.6062</u>
Target	Cyclegan	0.4010	1.2085

Table 2: Unsupervised Domain Adaptation w/Fine-Tuning(small samples)

Data	Method	mIoU	MAE(m)
Source	NA	0.6650	0.6538
Target	None	0.6885	0.6340
Target	HM	0.6811	0.5944
Target	FDA	0.6917	0.5839
Target	PDA	0.7014	0.5547
Target	LAB-HM	0.6862	<u>0.5627</u>
Target	Cyclegan	0.5714	0.8519

5 DISCUSSION

The selective-aligned data-based image transformation approach is able to outperform simple benchmarks using translated remotely sensed images from London to generate predictions with a model trained on Chicago for a multi-task urban canopy cover and height prediction algorithm in both zero-shot and few-sample fine-tune settings. Fine-tuning consistently outperforms zero-shot methods in our tests, however we still find the zero-shot methods useful as it may not always be practical to fine-tune an algorithm for every new target domain introduced.

Despite being able to align the distribution of London’s image pixels to be more similar to the Chicago images the algorithm was originally trained on, these methods cannot fully address the domain shift. This can arise from differences in geographical features, such as house shape, neighborhood layout, and building density specific to our target domain, London, which were not present

in the original image set, Chicago, as well as discrepancies in the conditions and instruments that were used to capture the imagery. Because of this, there seems to be a limit on how well these methods can perform zero-shot domain adaptation. In many scenarios however, there may not be viable ground truth estimates with which to fine-tune an algorithm on, or the algorithm may be operating within a very low-resource secure environment, such as a data-safe-haven, where any additional model training would be extremely time consuming. In these situations, FDA and PDA offered the best performance on canopy height and canopy cover predictions respectively, outperforming the baseline methods without the need to access the underlying model.

Surprisingly, nearly all of the fine-tuned algorithms outperformed the results the original algorithm acquired on the original image set. It is possible that by combining information from multiple locales, the total predictive power contained within the algorithm was increased. However, we did not have enough data to train a model from scratch on the London image set to compare against the fine-tuned model performance. Another possibility is that London is simply an easier setting for the task of canopy cover and height prediction than Chicago, as the buildings are much shorter on average.

While we tested these domain adaptation methods on an algorithm shifting from Chicago to London, future work would benefit from exploring domain shifts across a wider variety of geographic locales. Future work could also explore more advanced model-agnostic methods to improve the results in a zero-shot setting, test the methods over not only a one-to-one scenario but also many-to-many geographical scenarios, incorporate location information into the adaptation (Marsocci et al., 2023) or adaptation with large scale satellite image foundation models Bastani et al. (2023).

6 CONCLUSION

Overall, we find that selective-aligned simple image matching approaches can be an effective way to adapt a multi-task deep learning algorithm to a new geographic setting, especially when a small amount of fine-tuning is possible. Fourier domain adaptation and pixel distribution adaptation obtained the best results on our two tasks of canopy cover and canopy height prediction, although simple histogram matching techniques still provided moderate boosts over not using a domain adaptation method. Simple data-based domain adaptation approaches such as these provide easy, low-resource methods to enhance the utility of pre-trained models which utilize remote sensing imagery, when these algorithms are exposed to new, unseen domains. In many scenarios, fine-tuning or completely retraining an algorithm whenever it is exposed to a new domain can be either difficult to implement or prohibitively expensive. Researchers which are attempting to utilize pre-trained remote sensing algorithms should explore simple image matching techniques, before extensive and costly retraining, when applying these models to new geographic settings.

7 ACKNOWLEDGEMENTS

This work was supported by Towards Turing 2.0 under the EPSRC Grant EP/W037211/1, the Ecosystem Leadership Award under the EPSRC Grant EP/X03870X/1 & The Alan Turing Institute.

REFERENCES

- Fayyen Bastani, Piper Wolters, Ritwik Gupta, Joe Ferdinando, and Aniruddha Kembhavi. Satlaspretrain: A large-scale dataset for remote sensing image understanding. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 16772–16782, 2023.
- Alexander Buslaev, Vladimir I. Iglovikov, Eugene Khvedchenya, Alex Parinov, Mikhail Druzhinin, and Alexandr A. Kalinin. Albuementations: Fast and flexible image augmentations. *Information*, 11(2), 2020. ISSN 2078-2489. doi: 10.3390/info11020125. URL <https://www.mdpi.com/2078-2489/11/2/125>.
- John Francis, Mathias Disney, and Stephen Law. Monitoring canopy quality and improving equitable outcomes of urban tree planting using lidar and machine learning. *Urban Forestry Urban Greening*, 89:128115, 2023. ISSN 1618-8667. doi: <https://doi.org/10.1016/j.ufug>.

- 2023.128115. URL <https://www.sciencedirect.com/science/article/pii/S1618866723002868>.
- Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario March, and Victor Lempitsky. Domain-adversarial training of neural networks. *Journal of machine learning research*, 17(59):1–35, 2016.
- Valerio Marsocci, Nicolas Gonthier, Anatol Garioud, Simone Scardapane, and Clément Mallet. Geomultitasknet: remote sensing unsupervised domain adaptation using geographical coordinates. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2074–2084, 2023.
- Jean-Romain Roussel, David Auty, Nicholas C. Coops, Piotr Tompalski, Tristan R.H. Goodbody, Andrew Sánchez Meador, Jean-François Bourdon, Florian de Boissieu, and Alexis Achim. lidar: An r package for analysis of airborne laser scanning (als) data. *Remote Sensing of Environment*, 251:112061, 2020. ISSN 0034-4257. doi: <https://doi.org/10.1016/j.rse.2020.112061>. URL <https://www.sciencedirect.com/science/article/pii/S0034425720304314>.
- Emma Strubell, Ananya Ganesh, and Andrew McCallum. Energy and policy considerations for deep learning in nlp. *arXiv preprint arXiv:1906.02243*, 2019.
- Onur Tasar, SL Happy, Yuliya Tarabalka, and Pierre Alliez. Colormapgan: Unsupervised domain adaptation for semantic segmentation using color mapping generative adversarial networks. *IEEE Transactions on Geoscience and Remote Sensing*, 58(10):7178–7193, 2020.
- Devis Tuia, Claudio Persello, and Lorenzo Bruzzone. Domain adaptation for the classification of remote sensing data: An overview of recent advances. *IEEE Geoscience and Remote Sensing Magazine*, 4(2):41–57, June 2016. ISSN 2168-6831. doi: [10.1109/mgrs.2016.2548504](https://doi.org/10.1109/mgrs.2016.2548504). URL <http://dx.doi.org/10.1109/MGRS.2016.2548504>.
- Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell. Adversarial discriminative domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7167–7176, 2017.
- Ava Vali, Sara Comai, and Matteo Matteucci. Deep learning for land use and land cover classification based on hyperspectral and multispectral earth observation data: A review. *Remote Sensing*, 12(15), 2020. ISSN 2072-4292. doi: [10.3390/rs12152495](https://doi.org/10.3390/rs12152495). URL <https://www.mdpi.com/2072-4292/12/15/2495>.
- Stéfan van der Walt, Johannes L. Schönberger, Juan Nunez-Iglesias, François Boulogne, Joshua D. Warner, Neil Yager, Emmanuelle Gouillart, Tony Yu, and the scikit-image contributors. scikit-image: image processing in Python. *PeerJ*, 2:e453, 6 2014. ISSN 2167-8359. doi: [10.7717/peerj.453](https://doi.org/10.7717/peerj.453). URL <https://doi.org/10.7717/peerj.453>.
- Yanchao Yang and Stefano Soatto. Fda: Fourier domain adaptation for semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 4085–4095, 2020.
- Can Yaras, Kaleb Kassaw, Bohao Huang, Kyle Bradbury, and Jordan M Malof. Randomized histogram matching: A simple augmentation for unsupervised domain adaptation in overhead imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2023.
- Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pp. 2223–2232, 2017.
- Qing Zhu, Cheng Liao, Han Hu, Xiaoming Mei, and Haifeng Li. Map-net: Multiple attending path neural network for building footprint extraction from remote sensed imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 59(7):6169–6181, 2021. doi: [10.1109/TGRS.2020.3026051](https://doi.org/10.1109/TGRS.2020.3026051).
- Yang Zou, Zhiding Yu, BVK Kumar, and Jinsong Wang. Unsupervised domain adaptation for semantic segmentation via class-balanced self-training. In *Proceedings of the European conference on computer vision (ECCV)*, pp. 289–305, 2018.